

歌声の習熟度に関連する音響特徴量に関する基礎的検討 - 高周波域の特徴について -

A Fundamental Study of the Acoustic Features Related to the Classical Vocal Singing Proficiency: Features in the High Frequency Range.

寺内 大晴^{1*} 村山 諄² 山下 泰樹³ 舘 伸幸¹ 永井 孝⁴ 香山 瑞恵¹
Taisei Terauchi¹ Jun Murayama² Yasuki Yamashita³
Nobuyuki Tachi¹ Takashi Nagai⁴ Mizue Kayama¹

¹ 信州大学 工学部

¹ Faculty of Engineering, Shinshu University

² 信州大学大学院 総合理工学研究科

² Graduate School of Science & Technology, Shinshu University

³ 長野工科短期大学校 情報エレクトロニクス学科

³ Department of Information and Electronics, Nagano Prefectural Institute of Technology

⁴ ものつくり大学 技能工芸学部

⁴ Department of Manufacturing Technologists, Institute of Technologists

Abstract: The purpose of this study is to develop quantitative metrics based on acoustic features related to classical vocal singing proficiency. We have been exploring metrics that represent the quantitative proportion and sharpness of "singer's formants", frequency features specific to the singing voice. In this paper, we propose new metrics in the high frequency band of classical singing voice, and discuss the potential of the proposed metrics for singing voice evaluation.

1 はじめに

歌声の研究分野では、歌唱力と音響特徴量の関係について長年に渡って検討が重ねられてきた。本研究では、声楽歌唱技術の習熟度（以下、習熟度）と音響特徴量との関係性について焦点をあてているが、これについても多くの研究が存在している [1][2]。しかしながら、これらの先行研究の多くがオペラ歌手等のプロフェッショナルな歌手（以下、プロ歌手）を対象としており、声楽を学習する初學者の歌唱指導と音響特徴量の関係性については、未だ議論が不十分であると言える。また、初學者を対象とした歌唱指導においては、歌声の評価を指導者の主観にゆだねており、言語化が困難な指導となっている。我々は、歌声の周波数特性において特定の帯域に現れる Singer's formant と呼ばれる特徴 [1] に着目し、音響特徴量に基づく評価指標を提案、検証してきた [3]~[11]。

本論文では、2章で既存の歌声評価指標の詳細と関連する先行研究を示し、3章では提案評価指標とそれにつ

いて2つのリサーチクエスション（以下、RQ）を示す。4章ではRQに対して行った2つの分析における分析条件を述べ、5章では1つ目のRQに対して習熟度の異なる群間における提案評価指標の比較分析について考察し、6章では2つ目のRQに対して歌唱指導前後における提案評価指標の比較分析について考察する。

2 歌声評価指標に関する先行研究

本章では、本研究で用いる既存の歌声評価指標に関連する先行研究について述べる。本研究で用いる歌声評価指標としての音響特徴量は、Singer's formant とそれを含むとされる 2.4~4.0 kHz の帯域（以下、Singer's formant 帯域）について、その帯域の周波数成分（以下、Singer's formant 帯域成分）の割合と鋭さを定量化したものである。これらは、山辺ら及び佐久間らによって提案され [3]~[5]、歌唱の習熟度の評価に用いられてきた [3]~[11]。以下、これらの音響特徴量の算出法について述べる。

*20t2106g@shinshu-u.ac.jp

2.1 周波数成分の割合

歌声データに対する高速フーリエ変換 (Fast Fourier Transform: FFT) の結果に基づき, Singer's formant 帯域成分の割合を算出する. 歌声データの FFT から, 基本周波数と倍音がピークとして現れる. FFT の分析条件は, 標本化周波数:44.1 kHz, FFT ポイント数:2048 ポイント, 窓関数: ハニング窓とした.

音声波形の FFT の結果 $X(f)$ のパワースペクトル $P(f)$ を (1) 式で算出し, $P(f)$ から RMS 値を (2) 式で算出する. なお, RMS 値は, 量子化ビット数が 16 bit である音声波形の振幅スペクトルに対して, 振幅の最大値が 1 未満になるように変換した値である.

$$P(f) = 20 \log_{10} \frac{|X(f)|}{2^{16}} \quad (1)$$

$$RMS(f) = 10^{\frac{P(f)}{20}} \quad (2)$$

0 kHz~4.0 kHz までの帯域の周波数成分のうち, Singer's formant 帯域成分の割合を SFR (Singer's Formant Ratio) とし, (3) 式で算出する. 図 1 は SFR の考え方を図示したものである. 縦軸は RMS 値, 横軸は周波数をであり, グラフはプロ歌手による歌唱の例である. ここで, 青矢印が 0 kHz~4.0 kHz までの帯域, 赤矢印が Singer's formant 帯域を表す. SFR が大きいほど, Singer's formant 帯域に周波数成分が集中していることを表す.

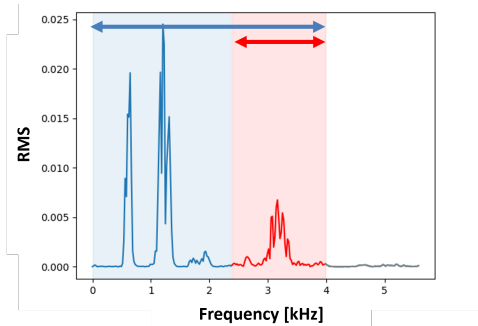


図 1: SFR の算出法

$$SFR = \frac{\sum_{f=2.4kHz}^{4.0kHz} RMS(f)}{\sum_{f=0kHz}^{4.0kHz} RMS(f)} \times 100 \quad (3)$$

2.2 周波数成分の鋭さ

本研究における周波数成分の強さとは, Singer's Formant 帯域成分が, 特定の周波数にどれだけ集中しているかを表す. これを評価するための指標として, 本節では, 声道の共振特性を表す LPC スペクトル包絡線 (以下, LPC 包絡線) と, LPC 包絡線から算出される Q 値, 2 凸について説明する.

2.2.1 LPC 包絡線

LPC 包絡線は以下の手順で求められる.

1. 音声波形にハニング窓をかける (以下, 波形 x)
2. LPC 次数 12, FFT のサンプル数 N , 自己相関関数のサイズ τ として, 波形 x の自己相関 r を (4) 式で求める

$$r(\tau) = \frac{1}{N-\tau} \sum_{k=0}^{N-1-\tau} x_k x_{k+\tau} \quad (\tau = 0, 1, \dots, \tau_{max}) \quad (4)$$

3. 自己相関 r を用いて, Levinson-Durbin アルゴリズムから LPC 係数を算出する

2.2.2 Q 値

図 2 は, 2.2.1 項に示した手順で求められた LPC 包絡線の例であり, 習熟度の高い歌唱者の場合, Singer's formant 帯域にピークが形成される. このピークの鋭さを Q 値とし, (5) 式で算出する. f_a は Singer's formant 帯域においてピークに相当する周波数を, f_b, f_c ($f_b > f_c$) はピークの両側に -3 dB となる周波数を指す. Q 値が大きいほど周波数成分が狭い範囲に集中していることを表す.

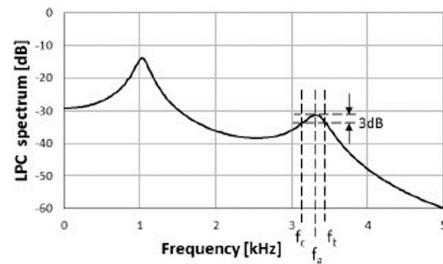


図 2: Q 値の算出法

$$Q = \frac{f_a}{f_b - f_c} \quad (5)$$

2.2.3 2 凸

音楽初学者の場合, Singer's formant 帯域成分が十分でなく, 図 2 のような Singer's formant 帯域のピークが形成されにくい. そのため Q 値の算出が困難であることがわかっている [5]. そこで, Singer's formant 帯域のピーク (LPC 包絡線における 2 個目のピーク) とそのピークの直前にある谷との差を 2 凸とし (図 3), これを Q 値に代わる指標とする.

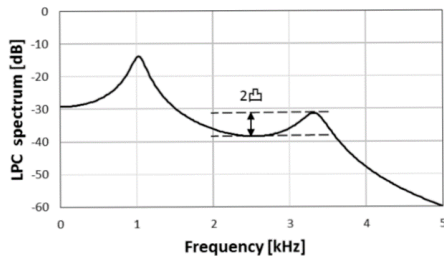


図 3: 2 凸の算出法

2.3 歌声評価指標と習熟度

2.1, 2.2 節に示した歌声評価指標は山下らの解析ツールで算出することができる [6]. この手法を使用した研究として、野田らは、学習者とプロ歌手の歌声分析から、これらの評価指標値の上昇が習熟度の向上に必要であることを示した [7]. また、吉田らは、学習者とプロ歌手が歌声評価指標によって区別可能であることを確認し、解析ツールによる結果とプロ歌手へのヒアリングに基づき、声楽における女声の「良い声」と「良くない声」を示す SFR, Q 値の閾値及び領域を仮定した [8][9]. さらに平井は、男声においてこれらの閾値及び領域を検討し、「良くない声」の領域を「非良領域」、「良い声」の領域を「良領域」、非良領域と良領域の間にある領域を「グレーゾーン」、良領域より値が大きな領域を「上側領域」と定義した (表 1) [10]. これらの先行研究によって、SFR, Q 値, 2 凸を用いて歌声の習熟度を定量的に評価できるようになった. これら以外の評価指標は、LPC 包絡線の形状 [11] や呼吸時における体表面の伸縮波形 [12] の観点が検討が継続されている.

3 研究目的

本章では、本研究で提案する歌声評価指標、本研究における RQ を述べる.

2.2.3 項において、初学者は Singer's formant 帯域成分が不十分なことから、Q 値の代わりに 2 凸が評価指標として用いられることを述べた. しかしながら、Singer's formant 帯域に LPC 包絡線の極大点が存在しない場合は、Q 値, 2 凸共に導出することができない. そのため、2 凸を用いても評価できない学習者に対して、習熟度が向上する過程を 0 でない値でフィードバックできる評価指標が必要である.

3.1 提案評価指標

前述の課題を解決するために、Singer's formant 帯域よりも高い周波数帯を対象とした新たな評価指標を提案する. ここでは、2.4~5.6 kHz の帯域成分のうち、

Singer's formant 帯域成分の割合を SFRH (SFR-High: 高周波数帯域における SFR) とし、(6) 式で算出する. 図 1 は SFRH の考え方を図示したものであり、青矢印が 2.4 kHz~5.6 kHz までの帯域、赤矢印が Singer's formant 帯域を表す. SFRH が大きいほど、高周波数帯域の中で Singer's formant 帯域に周波数成分が集中していることを表す. なお、Singer's formant 帯域の帯域幅と 4.0~5.6 kHz の帯域幅はユークリッド的に等しくなるように帯域幅を設定した. これにより、Singer's formant 帯域成分の合計と 4.0~5.6 kHz の帯域成分の合計が等しいとき SFRH は 50 になる. SFR と同様に Singer's formant 帯域成分が多いことが望ましいため、SFRH は 50 以上であることが望ましい.

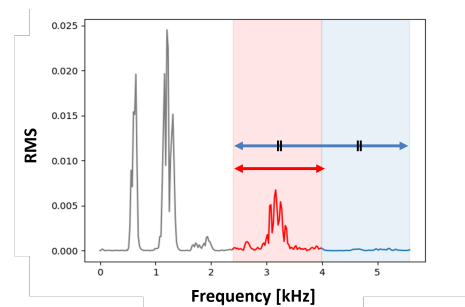


図 4: SFRH の算出法

$$SFRH = \frac{\sum_{f=2.4kHz}^{4.0kHz} RMS(f)}{\sum_{f=2.4kHz}^{5.6kHz} RMS(f)} \times 100 \quad (6)$$

3.2 RQ

本研究の目的は、Q 値や 2 凸を用いて歌声を評価できない学習者に対して、SFRH による歌声評価が可能かを検討することである. この目的を達成するために、以下の 2 つの RQ を設定した.

RQ.1 声楽学習者とプロ歌手とで、SFRH の分布に有意な差があるか

RQ.2 SFRH の上昇は学習者の習熟度の向上を表せるか

4 分析条件

本章では、本論文の 5 章と 6 章で行う分析に共通する分析対象楽曲、収音条件、被験者について述べる.

4.1 分析対象楽曲と歌声収音条件

分析に使用する楽曲は、「Caro mio ben」(作詞: 不明, 作曲: Tommaso Giordani) とした. 各被験者は図 5 の

表 1: 歌声評価指標の閾値と領域

\声の印象 (領域) 歌声評価指標\		良くない声 (非良領域)	(グレーゾーン)	良い声 (良領域)	(上側領域)
男声	SFR	0 ~ 20.3		21.8 ~ 30.0	
	Q 値	0 ~ 7.1		18.4 ~ 30.9	
女声	SFR	0 ~ 9.9		14.4 ~ 24.0	
	Q 値	0 ~ 5.6		18.0 ~ 34.7	

赤括弧で囲まれた区間を独唱する。收音は音の反響しない静かな部屋で行い、レコーダは LS-P2 (OLYMPUS) を使用した。被験者にはレコーダから 2 m 離れた正面立位で歌唱させた。音声はサンプリング周波数 44.1 kHz, 量子化ビット数 16 bit で收音した。分析対象とするデータは、声楽指導者が録音した。



図 5: Caro mio ben の譜面と歌唱区間

4.2 被験者のプロフィール

分析対象とした歌声は、オペラ歌手等のプロフェッショナルな歌手 (以下, プロ歌手/pro) の歌唱, 音楽大学声楽科在籍の学習者 (以下, mu), 教育学部在籍で音楽教育を専修する学習者 (以下, ed-A), ed-A とは異なる大学の教育学部在籍で音楽教育を専修する学習者 (以下, ed-B) の歌唱とした (表 2)。mu 及び ed-A, B の各被験者については, 各大学において歌唱指導を開始した直後の期間に收音した歌声データ (以下, 指導前データ) と, 歌唱指導を定期的に受けた後に收音した歌声データ (以下, 指導後データ) を分析対象とした。指導期間は, mu と ed-B が 8 か月, ed-A が 22 か月だった。本研究では, 時間の経過と習熟度の向上は一致していると仮定する。なお, 習熟度は pro, mu, ed-A, B (A, B は同程度) の順に高く, 以下, pro は「習熟度が高い群」, mu は「習熟度が中程度の群」, ed-A, B は「習熟度が低い群」とも表現する。

5 声楽の学習者とプロ歌手の比較

本章では, RQ.1 を検討する。歌唱区間から図 5 の赤下線部 “tanto” の/a/区間を抽出し, SFRH を算出した。その結果から, 学習者とプロ歌手の SFRH 分布を比較

被験者群\声種	男声	女声
pro	6	8
mu	2	8
ed-A	2	2
ed-B	0	5

することで, SFRH によって学習者とプロ歌手が区別可能かを考察する。

5.1 分析手法

各被験者の歌声データの/a/区間において算出した SFRH 分布の中央値を, 群毎に集約し比較した。なお, mu, ed-A, B については, 指導前データを使用した。

5.2 分析結果

図 6 に, 各群の SFRH の分布を箱ひげ図で表したものを示す。図中の橙実線は分布の中央値, 赤実線は SFRH が 50 の位置を表す。

Dunn-Bonferroni 法による多重比較 (有意水準 1%) の結果, (pro, ed-A), (pro, ed-B) において有意な差があった平均値±標準偏差は, pro: 93.2 ± 4.91, mu: 92.3 ± 2.90, ed-A: 71.3 ± 15.3, ed-B: 72.4 ± 15.1)。検定の結果, 有意な差があった群について, ブラケットを図 6 に付す。

5.3 考察

5.2 節の結果に基づき, RQ.1 に対して考察する。

pro と有意な差があった ed-A, B は, 4.2 節で述べたように, 分析対象の中で習熟度が低い群である。また, 習熟度が中程度である mu と, 習熟度が高い pro の間で有意な差が無かったことと, mu と ed-A, B は有意差がなかったことから, SFRH は習熟度が低い学習者か否かを区別可能な評価指標であることが示唆された。

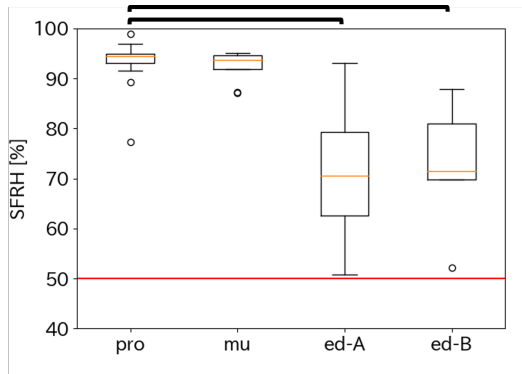


図 6: 各群の SFRH の分布と比較結果

6 学習者の指導前後の比較

本章では, RQ.2 を検討する. 歌唱区間から図 5 の赤下線部 “tanto” の/a/区間を抽出し, SFRH を算出した. その結果から, 学習者の指導前と指導後の SFRH 分布を比較することで, SFRH が習熟度の向上を表す評価指標として使用可能かを考察する.

6.1 分析手法

5 章の結果から, ed-A, B を分析対象とした. 各被験者の指導前後の歌声データの/a/区間において算出した/a/区間の SFRH 分布を比較した. また, 既存評価指標 (SFR, Q 値, 2 凸) による歌声評価と矛盾が無いかが検討するため, LPC 包絡線の時間変化の遷移を用いて Singer’s formant 帯域成分と高周波成分の増減を, 2 次元箱ひげ図 [13] で既存評価指標と SFRH の分布の変化を確認した.

6.2 分析結果

6.2.1 SFRH 分布の比較結果

図 7, 図 8 に, 各被験者の指導前後の SFRH 分布を箱ひげ図で表した. 点線枠で個人を区別し, 各点線枠内左が指導前, 右が指導後, 赤色実線は SFRH が 50 の位置を表す. また, 各被験者の指導後の SFRH 分布に対して, Mann-Whitney U 検定 (有意水準 1%) を行った結果を表 3 に示す. 検定の結果, 有意な差があった群について, 指導前から指導後にかけて SFRH が上昇したものを赤色, SFRH が下降したものを青色のブラケットで図 7, 図 8 に付す.

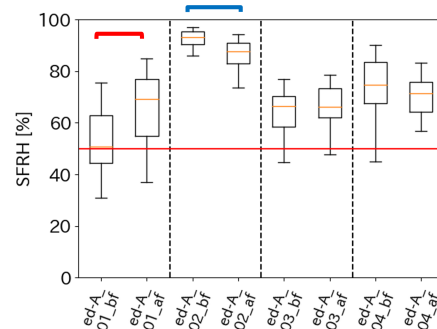


図 7: ed-A の SFRH 分布と比較結果

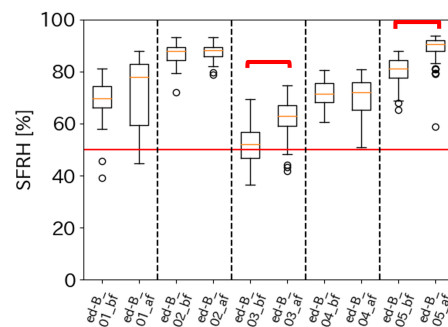


図 8: ed-B の SFRH 分布と比較結果

6.2.2 既存評価指標との対応の検証結果

まず, SFRH 分布が有意に上昇した 3 名の被験者 (ed-A.01, ed-B.03, ed-B.05) について検証する. 図 9 に ed-A.01 の検証結果を示す. 図の上段は LPC 包絡線の時間変化の遷移を表し, 左が指導前, 右が指導後, 実線枠内が Singer’s formant 帯域, 破線枠内が高周波域を表す. 下段は 2 次元箱ひげ図を表し, 横軸は左のグラフから SFR, Q 値, 2 凸, 縦軸はすべて SFRH である. 指導前が橙の点と赤の箱, 指導後が青の点と箱で描画されており, 箱は各データの第一四分位数と第三四分位数で表される頂点と, 中央値を表す箱の範囲を超過する線分で構成される. SFR, Q 値において, 良領域が桃色, 非良領域が灰色で描画されている. ed-A.01 について, LPC 包絡線の時間変化の遷移から Singer’s formant 帯域成分の増加が確認でき, 2 次元箱ひげ図から SFR, 2 凸の中央値の上昇, Q 値の四分位範囲の上昇方向への拡大が確認できる. ed-B.03 について, LPC 包絡線の時間変化の遷移から高周波成分の減少が確認でき, 2 次元箱ひげ図から Q 値, 2 凸の四分位範囲の上昇方向への拡大が確認できる (図 10). ed-B.05 について, LPC 包絡線の時間変化の遷移から Singer’s formant 帯域成分の増加が確認でき, 2 次元箱ひげ図から Q 値, 2 凸の四分位範囲の上昇方向への拡大が確認できる (図 11).

次に, SFRH 分布が有意に下降した 1 名の被験者 (ed-A.02) について検証する. ed-A.02 について, LPC 包

表 3: SFRH 分布の比較結果
(Mann-Whitney U 検定 (有意水準 1%))

被験者	p 値	平均値 ± 標準偏差	
		指導前	指導後
ed-A_01	***	53.0±11.8	66.1±12.6
ed-A_02	***	92.7± 2.7	86.4± 5.1
ed-A_03	.257	64.1± 2.7	66.4± 8.3
ed-A_04	.026	74.4±10.6	70.2± 6.8
ed-B_01	.087	69.0± 8.1	71.3±13.3
ed-B_02	.806	86.6± 4.1	87.3± 3.2
ed-B_03	***	52.2± 7.9	61.6± 7.8
ed-B_04	.196	71.6± 5.3	70.0± 7.8
ed-B_05	***	80.1± 5.1	88.3± 6.3

***: $p < .001$

絡線の時間変化の遷移から高周波成分の増加が確認でき、2次元箱ひげ図から SFR の中央値の上昇と Q 値、2凸の中央値の下降が確認できる (図 12)。

有意差が無かった5名の被験者については、大きく次の2種類に分類された。

1. SFR, Q 値, 2凸に変化なし (2名)
2. 指導前後共に、Q 値の中央値がグレーゾーン以上に存在する (2名)

これらに該当しない ed-B_01 については、SFR の上昇と Q 値、2凸の四分位範囲の上昇方向への拡大が確認された。この学習者は、中央値の上昇が見られたものの、検定では有意差がなかった学習者である。

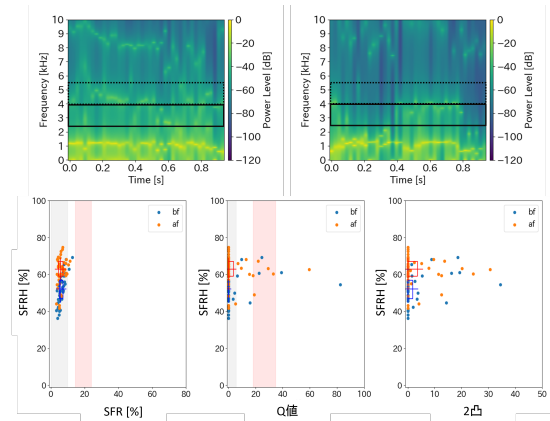


図 10: ed-B_03

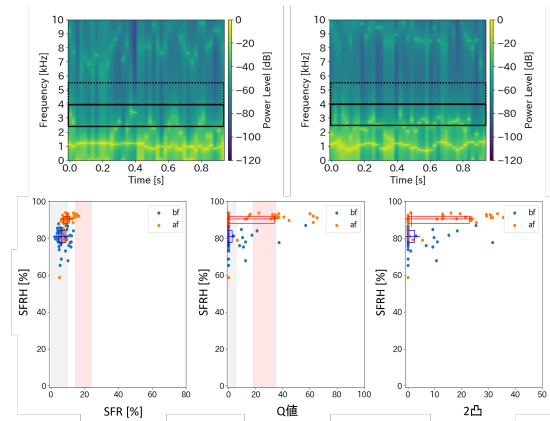


図 11: ed-B_05

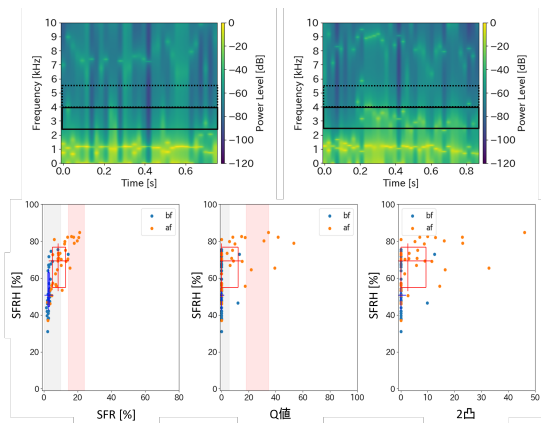


図 9: ed-A_01

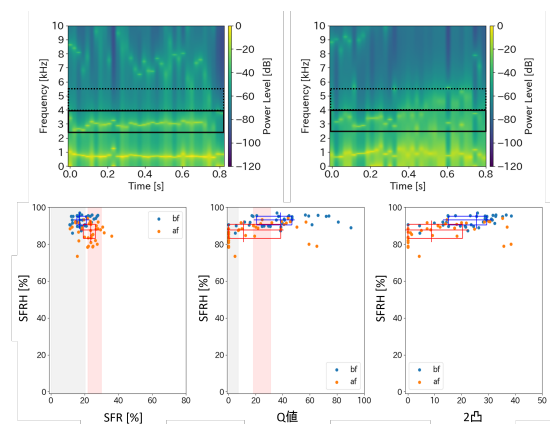


図 12: ed-A_02

6.3 考察

6.2 節の結果に基づき、RQ.2 に対して考察する。

SFRH が有意に上昇した3名について Q 値、2凸の

四分位範囲の拡大または中央値の上昇が確認されたこと、SFRH が有意に下降した 1 名について Q 値, 2 凸の下降方向への四分位範囲の拡大が確認されたことから、SFRH の変化は Q 値, 2 凸の変化と関連があることが示唆された。さらに、中央値による評価という観点から、Q 値, 2 凸で評価できない習熟度の向上が SFRH によって評価できることが示唆された。

以上のことから、SFRH の上昇は、Q 値, 2 凸に関連する歌声の習熟度の向上を表すことができ、Q 値, 2 凸によって評価できない学習者に対しても、SFRH は有効であることが示唆された。

7 おわりに

本稿では、学習者に対して Q 値, 2 凸では評価できない場合があることを示し、その解決策として SFRH を提案した。RQ.1 では、SFRH の分布が学習者とプロ歌手との間で有意な差があるか検証し、SFRH が習熟度の低い学習者か否かが区別可能であることを示した。RQ.2 では、SFRH の上昇が学習者の習熟度の向上を表せるか検証し、SFRH の上昇は、Q 値, 2 凸に関連する歌声の習熟度の向上を表せることを示した。以上のことから、SFRH は、Q 値, 2 凸に関連する歌声の習熟度を評価できる指標であり、特に Q 値, 2 凸では評価できない学習者に対しても 0 でない値で習熟度が向上する過程をフィードバック可能な指標であることが示唆された。

今後、SFRH による習熟度評価と歌声の聴感的印象の対応付け [14][15]、SFRH における「良い声」と「良くない声」の閾値の設定、既存評価指標と本提案指標の関連性の詳細な検証を行う。

謝辞

本研究は JSPS 科研費 22K00237 の助成を受けた。

参考文献

- [1] W.T.Bartholomew, “A Physical Definition of “Good Voice- Quality” in the Male Voice,” *J.Acoust. Soc. Am.*, Vol.6, No.1, pp.25-33 (1934)
- [2] J.Sundberg, “The KTH synthesis singing. Advances in Cognitive Psychology,” *Special issue on Music Performance*, Vol.2, No.2-3, pp.131-143 (2006)
- [3] 山辺大貴他, “歌声の心理的印象と音響特微量との対応付けによる歌唱の習熟度評価に関する基礎的検討”, 信学技報, Vol.112, No.266, pp.61-66 (2012)
- [4] 佐久間雄輝他, “歌声の習熟度と周波数特性との関係に関する考察”, 信学技報, Vol.113, No.349, pp.85-90 (2013)
- [5] 佐久間雄輝他, “歌声の習熟度に関連する周波数特性からみた音響特微量”, 信学技報, Vol.114, No.441, pp.45-50 (2015)
- [6] 山下泰樹他, “歌唱指導支援の検討に向けたデータベース環境の構築と解析提示方法”, 情処研報, Vol.2023-MUS-136, No.22, pp.1-7 (2023)
- [7] 野田美春他, “歌唱の習熟度に関連する周波数特性に基づく音居特微量を用いた学習者とプロ歌手の歌声評価”, 信学技報, Vol.115, No.444, pp.35-40 (2016)
- [8] 吉田祥他, “声楽の習熟度に関連する音響特微量に基づく歌声の評価方法に関する研究”, SIG-SKL-27-04, pp.13-20 (2019)
- [9] 吉田祥他, “声楽発声の数熟度に関連する音響特微量に基づく歌声の評価指標の提案”, 電子情報通信学会論文誌 D, Vol.J103-D, No.4, pp.247-260 (2020)
- [10] 平井雅人, “歌声の習熟度に関連する音響特微量とその時系列変化に関する研究”, 信州大学大学院総合理工学研究科修士論文 (未公刊), pp.46-56 (2022)
- [11] 矢川実紗他, “歌声の習熟度に関連するスペクトル包絡に基づく評価指標に関する基礎的検討”, 教育システム情報学会 2022 年度学生研究発表会 (北信越), pp.39-40 (2023)
- [12] 近藤由英他, “歌唱時における呼吸データの傾向と特徴”, 教育システム情報学会 2022 年度学生研究発表会 (北信越), pp.41-42 (2023)
- [13] 吉田祥他, “声楽の習熟度に関連する周波数特性に基づく音響特微量の特定個人の長期的変化”, 情報処理学会研究報告, Vol.2017-MUS-114, No.12 (2017)
- [14] 池田京子他, “歌声の「印象評価語」と「指導語」の抽出と分類 ~最適な指導語を求めて~”, 日本人工知能学会身体知研究会, Vol.33, pp.1-4 (2020)
- [15] 上村亮他, “成長を表す評価語の抽出と音響特微量からみる歌声の成長”, 教育システム情報学会 2020 年度学生研究発表会 (北信越), pp.39-40 (2021)